

# Comparing the Difference of Japanese College Students' English Listening Ability between the English and Japanese Answer Sheet Versions

Ryoko Muranaka

## I. Introduction

Test techniques are a means of eliciting behavior from students which can be a reliable and valid indicator of their ability and also which can be reliably scored. Multiple-choice testing is one of them. Students must identify or select the correct or most appropriate options.

Where a multiple-choice format is used to test listening comprehension, there seems to be a potential for disparity in the manner in which the answers are presented, i.e., English or the students' native language.

The specific question addressed in this research concerns the testing of oral communication in English as a foreign language among Japanese college students stated as a null hypothesis. That is ; *There is no difference in the results when Japanese college students take multiple-choice English listening tests in which answer options to questions are written in Japanese than those who take the same test with answer options to questions written in English.*

## **II. Purpose**

The main purpose of this study is to compare the difference of students' English listening comprehension between the English and Japanese answer sheet versions.

## **III. Method**

### 1. Subjects

The subjects in this study were 202 native Japanese University students majoring in Business Administration. Nearly all of these students have been studying English in Japan for more than six years, and most were from eighteen to twenty-one years of age.

### 2. Test Instruments and Procedures

The test instrument utilized in this research was adapted from the multiple choice design of TOEFL. and, more specifically, adapted or arranged from units 6, 8, 10, 11, 14 and 15 from "Listening for TOEFL" edited by Trokeloshivili, S. Davis and N. Sakikawa (1996). The TOEFL listening test format facilitates this research paradigm since only the multiple choice options are written on the answer sheet.

The students responded to ten items with four multiple-choice options each written either in English or in Japanese. The English listening test items were presented by means of tape-recording of native speakers' dialogue.

### 3. Data Analysis

The data analysis was designed to verify the research question by determining central tendency (frequency and means), validity (comparison of means-t test) and item statistics (chi-square).

## IV. Results and Discussion

### 1. Testing Results

*Distribution and Variability.* The results of the testing are shown in Tables 1 and 2 below.

**Table 1 Distribution of Test Results Totals**

Score	Frequency		Percent	
	Japanese	English	Japanese	English
3	1	1	1	1
4	0	4	0	3.8
5	3	5	3.1	4.8
6	5	15	5.1	14.4
7	15	21	15.3	20.2
8	38	33	38.8	31.7
9	32	22	32.7	21.2
10	4	3	4.1	2.9
Total	98	104	100	100

**Table 2 Variability of Results Means**

	N	Mean	S.D.	F	sig.	t	d.f.	sig.
Japanese	98	8.01	1.17	9.22	<.05	3.11	196	<.05
English	104	7.43	1.44					

These data show a small but significant difference in the mean scores of the two groups. The modes of the two groups are identical.

*Item Analysis.* The correct/incorrect responses to each item by groups are shown in Table 3. Mean scores appear in Tables 4 and 5. The summary of individual chi-square comparisons are in Table 6.

**Table 3 Correct and Incorrect Test Item Responses**

Item	1		2		3		4		5	
Group	J	E	J	E	J	E	J	E	J	E
Incorrect (N)	20	23	3	2	21	15	20	31	14	39
Correct (N)	78	81	95	102	77	89	78	73	84	65
Total (N)	98	104	98	104	98	104	98	104	98	104
Incorrect (%)	20	22	3	2	21	14	20	30	14	38
Correct (%)	78	78	97	98	79	86	78	70	86	62
Total (%)	100	100	100	100	100	100	100	100	100	100

Item	6		7		8		9		10	
Group	J	E	J	E	J	E	J	E	J	E
Incorrect (N)	2	16	77	76	3	15	29	37	6	13
Correct (N)	96	88	21	28	95	89	69	67	92	91
Total (N)	98	104	98	104	98	104	98	104	98	104
Incorrect (%)	2	15	79	73	3	14	30	36	6	13
Correct (%)	98	85	21	27	97	86	70	64	94	87
Total (%)	100	100	100	100	100	100	100	100	100	100

Note : J stands for Japanese version answers and E stands for English version answers.

**Table 4 Test Item Difficulty Mean and S.D. (Japanese)**

Item	Mean	S.D.	most difficult
7	0.21	0.41	
9	0.70	0.46	
3	0.79	0.41	
4	0.80	0.41	
1	0.80	0.41	
5	0.86	0.35	
10	0.94	0.24	
8	0.97	0.17	
2	0.97	0.17	
6	0.98	0.14	Easiest

**Table 5 Test Item Difficulty Mean and S.D. (English)**

Item	Mean	S.D.	most difficult
7	0.27	0.45	
5	0.63	0.49	
9	0.64	0.48	
4	0.70	0.46	
1	0.78	0.42	
6	0.85	0.36	
8	0.86	0.35	
3	0.86	0.35	
10	0.87	0.33	
2	0.98	0.14	easiest

**Table 6 Test Item Chi-Square Comparison**

Item	1	2	3	4	5	6	7	8	9	10
sig.	ns	ns	ns	ns	p<.001	p<.001	ns	p<.01	ns	ns

Note. The data for individual items appear in Appendix C.

## 2. Discussion

Table 1 shows that in the Japanese version most of the students received a score of 7, 8 or 9. The score with the highest percentage of students was 8. This table also shows that in the English version most of the students received a score of 6, 7, 8 or 9 with the majority receiving a score of 8. In terms of the score modes, these two groups are identical. In other words, there seems to be a similarity in the shape of the two frequency distributions.

There is a significant difference shown in Table 2 at the .05 level in the means of the two groups. That is, those who take the Japanese version where the multiple-choice options are given in Japanese and those who take the English version where the multiple-choice options are given in English. As a whole, we can say with great confidence that the students who are given the multiple-choice items in Japanese do better on the test than those who are given the options in English. Our research question, stated in the null form, *There is no difference in the results when Japanese college students take multiple-choice English listening tests in which answer options to questions are written in Japanese than those who take the same test with answer options to questions written in English*, has not been supported.

The results of the item difficulty analysis, Table 3, lists the proportion of the number of people who got each item correct. Tables 4 and 5 demonstrate the same thing using Flanagan's item difficulty indices.

In the Japanese version, item 7 is the most difficult and item 6 is the easiest. However, item 7 is by far the most difficult, judging from scores observed in the tables. In the English version, item 7 is the most difficult and item 2 is the easiest. And again item 7 is far more difficult than the others. Between the Japanese version and the English version, the most difficult item is common to both tests (English version and Japanese version), while the easiest one is different. As a whole, the order of difficulty of each item in each version of the test (English and Japanese) is different, and the students seemed to have performed differently depending on what language was used on the question sheet.

Table 6 indicates that there has been a significant difference in the chi-square distributions for items 5, 6 and 8 of the two groups. Individual chi-square results appear in Appendix C. Another way of looking at these items is that the difference between the students who got items correct and the students who got items incorrect in the Japanese version is statistically bigger than that in the English version. In other words, for items 5, 6 and 8, students do better in the Japanese version than in the English version. The result of this present research has successfully rejected the null research question, because students do better in these items in the Japanese version than in the English version. Although there are 7 other items where there is no statistical difference in the distribution between the two groups, we still can say with great confidence that students do better in the Japanese version test than in the English version test. On the basis of these data analyses included in Tables 1 through 6, we could say that our research question, *There is no difference in the results when Japa-*

*nese college students take multiple-choice English listening tests in which answer options to questions are written in Japanese than those who take the same test with answer options to questions written in English, has not been supported.*

The present researcher acknowledges the failure to do a try-out and discrimination analysis. Those procedures could have led to the removal or re-write of items with low levels of discrimination and potential greater effectiveness of the research. Since no try-out was conducted, some of the items which had low discriminating power contributed very little to the item analysis.

The two language levels and their translations seemed to be equivalent in the professional opinions of the present researcher and the translation reviewer ; that is, each item in both languages used equivalent levels of grammatical construction, vocabulary, syntax etc. Similarly, sentence structures in both versions seemed to be simple enough for students to understand. However, it might be possible to say that the lack of students' background knowledge or word connotations in three items 5, 6 and 8 influenced their weaker performance in the English version. For example, if, in item 5, students are not familiar with or not fond of golfing, it could be a little difficult to understand the situation. In item 6, if students were not aware of the language paraphrasing "from open to close" this might have confused the students in the English version. In item 8, the word "direct" in English might not have been familiar to students. Thus, one possible reason for the score difference between the two versions in at least these three items is the lacking of students' background knowledge in English. It might be that students could concentrate more on listening

rather than trying to understand the meaning of difficult English sentences or words when the choices were given in their native language.

Although it is difficult to identify an outside reason for the difference in the results, the aforementioned explanations for three items possibly identify some language-oriented reasons for the resulting variance.

## **V. Conclusions**

The research question has not been supported as follows :

- 1) The mean score difference of the listening test scores between the two versions (English and Japanese) suggests that students taking the Japanese version do better than those taking the English version.
- 2) The Chi-square test results show that, for three items, students taking the Japanese version do significantly better than those taking the English version. Accordingly, even in the details, the basic premise of the research question has not been supported.

The present research shows that, in the English listening comprehension test, students perform measurably different when the options are given in their mother tongue probably because they choose their answers without having difficulty in understanding the language problem of option choices. This seems to permit students to focus more on listening rather than reading the choices and understanding the meaning of the English sentences.

It is commonly agreed that foreign language education should be done in the target language, so the foreign language test should also



be done in that target language. However, the present research does not necessarily support that general testing methodology, especially when it comes to a listening test.

## APPENDICES

### Appendix A (English Version)

1. (woman) The cooking class is canceled this week.  
 (man) Oh, no. I'm going away on a business trip next week, and this is my only free week.  
 (narrator) What is the man's problem?  
 (A) He loves to cook.  
 (B) He can't go to cooking classes very often.  
 (C) He is not very good at cooking.  
 (D) He can't come to the class next week.
2. (woman) I find that jogging in the morning is very refreshing.  
 (man) I prefer jogging during the evening.  
 (narrator) Why does the woman jog in the morning?  
 (A) Because it makes her feel better.  
 (B) Because she's scared to jog during the night.  
 (C) Because she likes getting up early.  
 (D) Because it makes her tired.
3. (man) I need some help in the office this summer. Are you interested?  
 (woman) Do I need a college degree?  
 (narrator) What is the woman worried about?  
 (A) If she can get the job.  
 (B) If she had to graduate from college before starting the

summer job.

(C) If she was going away during the summer.

(D) If the man was going to change his mind.

4. (woman) I've been waiting for the machinery for five months.

(man) Two or three months is normal, but that's ridiculous.

(narrator) What is the man saying about the delay?

(A) It's scary.

(B) The woman should wait longer.

(C) The delay is too long.

(D) The machinery is broken by now.

5. (woman) I played 7 holes of golf this morning. Then it started to rain, but I continued.

(man) I really don't care much for golf.

(narrator) What did the woman do?

(A) She went home.

(B) She carried on playing even though it started to rain.

(C) She stopped playing.

(D) She played 7 holes.

6. (man) I propose you keep the shop open during the Christmas holidays.

(woman) I think that's a good idea.

(narrator) What is the man saying?

(A) He thinks the shop should be closed.

(B) He wants to go home for Christmas.

(C) He doesn't want the shop to close at Christmas time.

(D) He wants to make money.

7. (man) I would like to conclude this meeting by saying that it's been

a pleasure working with you.

(woman) We do work very well together.

(narrator) "Conclude" probably means :

(A) Begin.

(B) Hesitate.

(C) Finish.

(D) Celebrate.

8. (woman) Could you direct me to the nearest post office?

(man) I'm sorry, I don't know where it is.

(narrator) What is the woman asking?

(A) How to get to a post office.

(B) How much a stamp costs.

(C) Show her what a post office looks like.

(D) Direct her to the farthest post office.

9. (woman) Have you watered the plants yet?

(man) I thought you were going to do it.

(narrator) What is the man implying?

(A) That he hasn't watered the plants yet.

(B) That the woman should have watered the plants.

(C) That the woman hasn't watered the plants.

(D) That the plants don't like water.

10. (man) I have to hand in my schedule for my summer vacation next week.

(woman) I'm postponing mine until October because I'm too busy now.

(narrator) When will the woman take her vacation?

(A) Next week.

- (B) Next year.
- (C) In October.
- (D) In the summer.

## Appendix B

### *Ten Multiple-Choice Options in the Japanese Translation*

1. (A) 彼は料理が大好きだ。  
(B) 彼は頻繁には料理教室には行けない。  
(C) 彼は料理があまり得意ではない。  
(D) 彼は来週はその教室へは来れない。
2. (A) それによって彼女は気分がよくなるから。  
(B) 彼女は夜はジョギングするのが怖いから。  
(C) 彼女は早起きするのが好きだから。  
(D) それによって彼女は疲れるから。
3. (A) 彼女は仕事を得られるかどうか。  
(B) 夏の仕事を始める前に彼女は大学を卒業しなければならなかったか。  
(C) 彼女は夏の間は出かけていたかどうか。  
(D) その男性が心変わりをしてしまったかどうか。
4. (A) それは恐ろしい。  
(B) 女性はもう少々待たねばならない。  
(C) その遅れが少々長すぎる。  
(D) その機械はもう今頃は壊れてしまっている。
5. (A) 彼女は帰宅した。  
(B) 雨が降りだしたが彼女はプレーを続けた。  
(C) 彼女はプレーを止めた。  
(D) 彼女はプレーを7ホールした。

6. (A) 彼は店を閉めるべきだと思う。  
(B) 彼はクリスマスには帰宅したいと思う。  
(C) 彼はクリスマスの時期には店を閉めたくないと思う。  
(D) 彼は金もうけがしたい。
7. (A) 始める。  
(B) ためらう。  
(C) 終える。  
(D) 祝う。
8. (A) 郵便局へはどのように行くか。  
(B) 切手はいくらか。  
(C) 彼女に郵便局がどんな所か教える。  
(D) 一番遠い郵便局に彼女を案内する。
9. (A) 彼がまだ植木に水をやっていないこと。  
(B) その女性が植木に水をやるべきだったこと。  
(C) その女性が植木に水をやらなかったこと。  
(D) その植木は水を好まないこと。
10. (A) 来週。  
(B) 来年。  
(C) 十月に。  
(D) 夏に。

**Appendix C***The Chi-Square Test Results of the Distribution of the Four Categories in Each Item***item 1**

	J-version	E-version	Total	F-value	df	sig
incorrect (%)	20	23	43	0.088	1	ns
	20.4	22.1	21.3			
correct (%)	78	81	159			
	79.6	77.9	78.7			
Total (%)	98	104	202			
	100	100	100			

**item 2**

	J-version	E-version	Total	F-value	df	sig
incorrect (%)	3	2	5	0.271	1	ns
	3.1	1.9	2.5			
correct (%)	95	102	197			
	96.9	98.1	97.5			
Total (%)	98	104	202			
	100	100	100			

**item 3**

	J-version	E-version	Total	F-value	df	sig
incorrect (%)	21	15	36	1.691	1	ns
	21.4	14.4	17.8			
correct (%)	77	89	166			
	78.6	85.6	82.2			
Total (%)	98	104	202			
	100	100	100			

**item 4**

	J-version	E-version	Total	F-value	df	sig
incorrect (%)	20	31	51	2.362	1	ns
	20.4	29.8	25.2			
correct (%)	78	73	151			
	79.6	70.2	74.8			
Total (%)	98	104	202			
	100	100	100			

**item 5**

	J-version	E-version	Total	F-value	df	sig
incorrect (%)	14	39	53	14.049	1	p<.001
	14.3	37.5	26.2			
correct (%)	84	65	149			
	85.7	62.5	73.8			
Total (%)	98	104	202			
	100	100	100			

**item 6**

	J-version	E-version	Total	F-value	df	sig
incorrect (%)	2	16	18	11.068	1	p<.001
	2	15.4	8.9			
correct (%)	96	88	184			
	98	84.6	91.1			
Total (%)	98	104	202			
	100	100	100			

**item 7**

	J-version	E-version	Total	F-value	df	sig
incorrect (%)	77	76	153	0.829	1	ns
	78.6	73.1	75.7			
correct (%)	21	28	49			
	21.4	26.9	24.3			
Total (%)	98	104	202			
	100	100	100			

**item 8**

	J-version	E-version	Total	F-value	df	sig
incorrect (%)	3 3.1	15 14.4	18 8.9	8.025	1	p<.01
correct (%)	95 96.9	89 85.6	184 91.1			
Total (%)	98 100	104 100	202 100			

**item 9**

	J-version	E-version	Total	F-value	df	sig
incorrect (%)	29 29.6	37 35.6	66 32.7	0.822	1	ns
correct (%)	69 70.4	67 64.4	136 67.3			
Total (%)	98 100	104 100	202 100			

**item 10**

	J-version	E-version	Total	F-value	df	sig
incorrect (%)	6 6.1	13 12.5	19 9.4	2.408	1	ns
correct (%)	92 93.9	91 87.5	183 60.6			
Total (%)	98 100	104 100	202 100			

**BIBLIOGRAPHY**

- Alderson, J.C., Clapham, C. and Wall, D. (1995). *Language Test Construction and Evaluation*. Cambridge : Cambridge University Press.
- Allison, D. (1999). *Language Testing and Evaluation : An Introductory Course*. Singapore : Singapore University Press.
- Bachman, L.F. (1995). *Fundamental Considerations in Language Testing*. Oxford : Oxford University Press.



- Bachman, L.F. and Palmer, A.S. (1996). *Language Testing in Practice*. Oxford : Oxford University Press.
- Brown, H.D. (1994). *Teaching by Principles : An Interactive Approach to Language Pedagogy*. New Jersey : Prentice Hall Regents.
- Brown, H.D. (2004). *Language Assessment—Principles and Classroom Practices*. New York : Pearson Education.
- Brown, J.D. (1996). *Testing in Language Programs*. New Jersey : Prentice Hall Regents.
- Brown, J.D. (Ed.). (1998). *New Ways of Classroom Assessment*. Alexandria, VA : TESOL.
- Buck, G. (1988). Testing Listening Comprehension in Japanese University Entrance Examinations. *JALT Journal* 10, 1 & 2, 15–42.
- Buck, G. (1994). The appropriacy of psychometric measurement models for testing second language listening comprehension. *Language Testing* 11, 2, 145–170.
- Buck, G. (2001). *Assessing Listening*. Cambridge, U.K. : Cambridge University Press.
- Davies, A. (1984). Validating three tests of English language proficiency. *Language Testing* 1, 1, 50–69.
- Douglas, D. (2003). *English Language Testing in U.S. Colleges and Universities*. Washington, D.C. : NAFSA : Association of International Educators.
- Davies, A., Brown, A., Elder, C., Hill, K., Lumley, T., McNamara, T. (Eds.). (1999). *Dictionary of Language Testing*. Cambridge : Cambridge University Press.
- Dunkel, P. (1991). Listening in the native and second/foreign language : Toward an integration of research and practice. *TESOL Quarterly* 25, 3, 431–457.
- Flowerdew, J. (Ed.). (1994). *Academic Listening : Research Perspectives*. Cambridge : Cambridge University Press.
- Henning, G. (1987). *A Guide to Language Testing : Development, Evaluation, Research*. Cambridge : Newbury House Publishers.
- Hughes, A. (2003). *Testing for Language Teachers*. Second Edition. Cambridge : Cambridge University Press.
- Linn, R.L. (1993). *Educational Measurement* (3 rd ed.). Phoenix, AZ : The Oryx Press.
- Osterlind, S.J. (1998). *Constructing Test Items : Multiple–Choice, Constructed–Response, Performance, and Other Formats* (2 nd ed.). Boston : Kluwer Academic Publishers.

- Richards, J.C. (1983). Listening comprehension : Approach, design, procedure. *TESOL Quarterly* 17, 2, 219–239
- Rost, M. (1996). *Listening in Language Learning*. New York : Longman Inc.
- Trokeloshvili, D., Davies, S. and Sakikawa, N. (Eds.). (1996). *Listening for TOEFL*. Tokyo : Eicho-sha.
- Weir, C.J. (1990). *Communicative Language Testing*. New York : Prentice Hall.
- Weir, C.J. (1993). *Understanding and Developing Language Tests*. London : Prentice Hall.
- Weir, C.J. (2005). *Language Testing and Validation : An Evidence-Based Approach*. New York : PALGRAVE MACMILLAN.